TECHNICAL APPENDIX: CHAPTER X

ECONOMIC EFFECTS SECTION

Principal Components Analysis

For nonmetropolitan counties in the contiguous United States (n=2260), principal components analysis was performed on an identical set of resource variables for each of the three regions, North, South, and West (Table 1). Principal components analysis begins by creating a correlation matrix of the variables. Highly correlated variables are linearly combined into a new set of variables called 'components'. The first principle component ascribes a line of bet fit in geometric space across all of the data. It accounts for as much of the total variance in the correlation matrix across all variables as possible. The first principle component is the best summary of the entire data set. An eigenvalue, a summary measure of the amount of variance in the correlation matrix explained by the factor, is calculated as well.

Once the first component is determined, the explained variance is removed from the correlation matrix and the second component is calculated. This component is geometrically orthogonal to the first line, and attempts to explain as much of the remaining variance in the correlation matrix as possible. A third component orthogonal to the first two is calculated next. The process continues until all of the variation in the correlation matrix is explained. When complete, the principal components are independent of each other and provide a concise, comprehensive description of the original data.

An eigenvector (a column of weighted scores) decomposes the contribution of each of the original variables to the principal component. The eigenvector is transformed into a principal component loading score by multiplying each element of the vector by the square root of the eigenvalue. Variables with loading scores above 0.5 are generally considered to contribute to the meaning of the principal component.

For each region, resources were classified into major types, depending on whether they were <u>urban</u> and developed resources, such as golf courses, swimming pools, etc., dispersed <u>land</u> resources such as forest land, wilderness acres, or state park land, water resources, including rivers or lakes, or winter resources. Separate principal components analyses were performed for each class of resource, Urban, Land, Water, and Winter. In addition, analyses were done separately for each region, since it was not assumed that resources would correlate identically in different regions Factors with eigenvalues greater than 1.0 were retained. In each region, 16 factors were retained, four Urban, six Land, four Water, and two Winter[1]. Groupings of factors were fairly similar across regions. In general, the retained factors accounted for between 55 percent and 68 percent of the variation in the resource variables.

Cluster Analysis

Cluster analysis is used to generate classifications of similar observations. Resulting clusters are intended to minimize variation within groups while maximizing differences between groups. The clustering method used was a hierarchical algorithm known as Ward's Minimum Variance Method. A point representing each county's variable scores (for variables used to cluster) is set in geometric space. Using a generalized Pythagorean Theorem, Ward's Method calculated the distance between each point. The closest points are joined, and replaced by a new point, representing the combination of the two individual points.

At each step, the error sum of squares is calculated to measure the amount of information loss associated with the joining of points. By identifying where large jumps in the error term occur, it is possible to identify the boundary between relatively distinct clusters. The resulting clusters can be mapped to display the dispersion of similar counties or to identify larger regions.

Counties were clustered according to values on five variables that described the type of county.

---

[1] Factor score tables, cluster maps, and tabular results from cluster and regression analyses in this section are available upon request from Dr. Donald B.K. English, Forestry Sciences Lab, 320 Green St., Athens, GA 30602.

The variables included population density, and the proportion of county acres in cropland, forests, pasture/rangeland, and mountains. The data source was the 1997 NORSIS dataset, described in detail elsewhere in this Assessment. In each region, eight clusters were retained. In each region, these eight clusters accounted for about 86 percent of the total variation in the clustering variables. Figures 1 through 6 show the counties in each cluster for each of the three regions.

Minimum Requirements

The minimum requirements technique established export shares (the portion of economic activity in a sector that is due to nonresident demand) for each county, by sector. In each cluster and for each sector, the minimum nonzero proportion of economic activity is identified. It is assumed that in this minimum ratio county, all production is needed to meet local demand, i.e., there is no ‘export’ of those services to support demand by nonresidents of the county. In all other counties in the cluster, the excess above that minimum ratio is assumed to meet export (out of county) demand. Employment and income data for each nonmetropolitan county, and for each of the four sectors examined (Hotels and other lodging, Eating and Drinking places, Recreation and Amusement Services, and Other retail trade), came from sectoral data in the 1993 Micro-Implan data set, developed by the MicroImplan Group at the University of Minnesota. The calculation to determine the export share for county *I* and sector *j* is:

$$ES_{ij} = \left[\frac{ec_{ij}}{ec_{iT}} - \min\left(\frac{ec_{ij}}{ec_{iT}}\right)\right] ec_{iT}$$

where:

Es_{ij} = export share for county I and sector j ;

ec_{ij} = economic activity (employment or income) for sector j in county I;

ec_{iT} = economic activity for county I, summed over all sectors;

min(.) = minimum function, identifying the minimum nonzero value across all counties in the cluster for county I.

Regression Analysis

For counties whose export activity was greater than zero, it was necessary to estimate the proportion that was due to recreation activity. Clearly, any county that had zero export activity for a sector had zero recreation-related activity. Population was assumed to be a proxy for the amount of business and family-related travel that occurred in a county. The principal component scores calculated in the first step served a independent variables describing the level of different types of recreation resources in the county. Both linear and log-linear specifications were examined. The log-linear form had the conceptual advantage that predicted values would all be positive. It turned out to provide superior fit to the data, as well. Therefore, the model estimated in the regressions was:

$$\log(ES_{ij}) = \alpha_j + \beta_{POPj} POP_i + \beta_{RECj} REC_i$$

where:

POP_i = Population of county I;

REC = Vector of principal components scores for county I;

α_j , β_{POPj} , β_{RECj} = parameters to be estimated for sector j .

Equations were estimated separately for each sector in each region. From the results, recreation-related economic activity for each sector in each county (ER_{ij}) was calculated by removing the effect of population-related activity. The specific calculation was:

$$ER_{ij} = \exp(\hat{\alpha}_j + \hat{\beta}_{RECj} REC_i)$$

Technical Appendix Table X.1–Definitions of variables used in principal components analysis

Urban facilities:

parks & recreation departments
tour operators + sightseeing tour operators
playgrounds + # recreation centers
private +public swimming pools
private + public tennis courts
organized camps
tourist attractions + # historical places
amusement places
fairgrounds
local or county parks
private + public golf courses
ISTEA funded greenway trails
Estimate of acres of urban and built up land from 1995 National Resources Inventory (NRI)

Land resources:

guides services
hunting/Fishing preserves, clubs, lodges
BLM public domain acres
Acres of mountains
Acres of cropland, pasture, rangeland
USDA-FS National Forest and Grassland acres
FWS refuge acres open for recreation
WOODALLS # private campground sites
WOODALLS # public campground sites
NPS federal acres
NRI estimate of forest acres
Acres managed by Bureau of Reclamation, Tennessee Valley Authority, Corps of Engineers
Total rail-trail miles
State Park acres
The Nature Conservancy acres with public access
National Wilderness Preservation System acreage: Total 1993

Technical Appendix Table X.1 (Continued.)

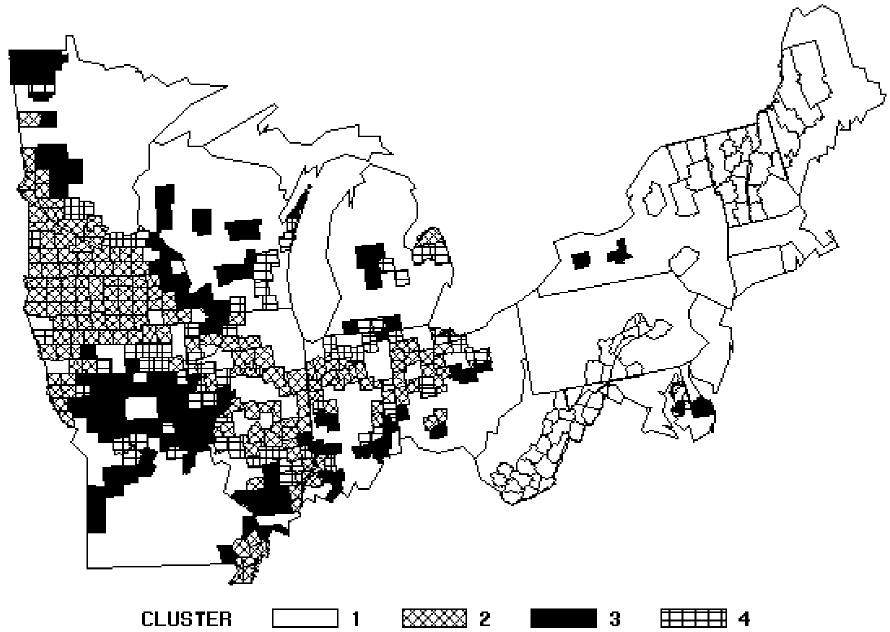
Water resources:

Marinas
canoe outfitters + rental firms+ raft trip firms
diving instruction or tours + snorkel outfitters
guides services
fish camps + private or public fish lakes, piers, ponds
American Whitewater Association total whitewater river miles
Wild & Scenic River miles: Total 1993
NRI water 2-40,<2 ac, >= 40 ac (lake or reservoir.)
NRI stream<66' wide + 66-660' wide + >=1/8 mi wide
NRI water body >= 40 ac. (bay, gulf, or estuary)
NRI wetland acres
Nationwide Rivers Inventory Total river miles, any outstanding value

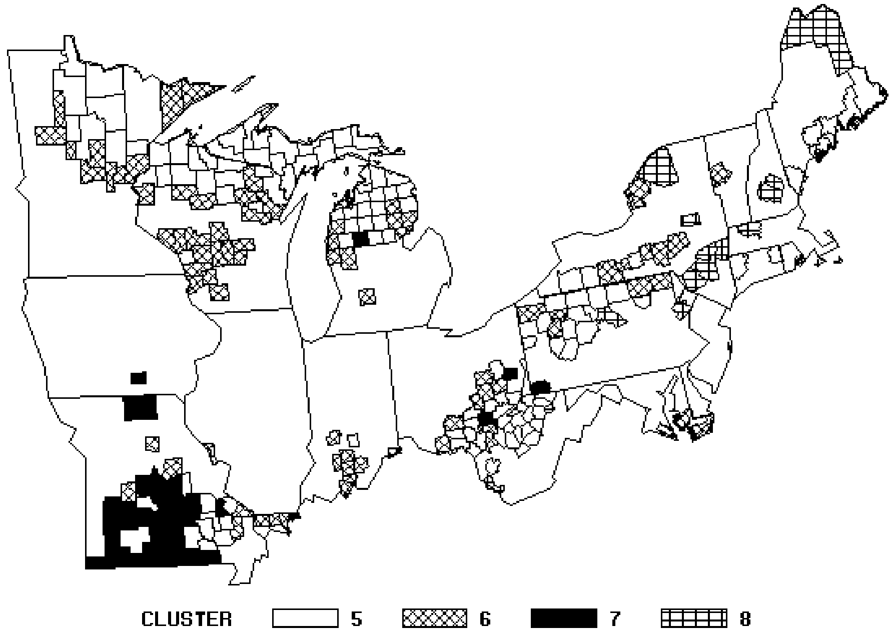
Winter resources:

Cross-Country Ski Areas Association # XCski firms + public XC centers
International Ski Service Skiable acreage
Federal land acres in counties with > 24" annual snowfall
Agricultural acres in counties with >24" annual snowfall
Acres of mountains in counties with >24 " annual snowfall
Acres of forestland in counties with >24" annual snowfall

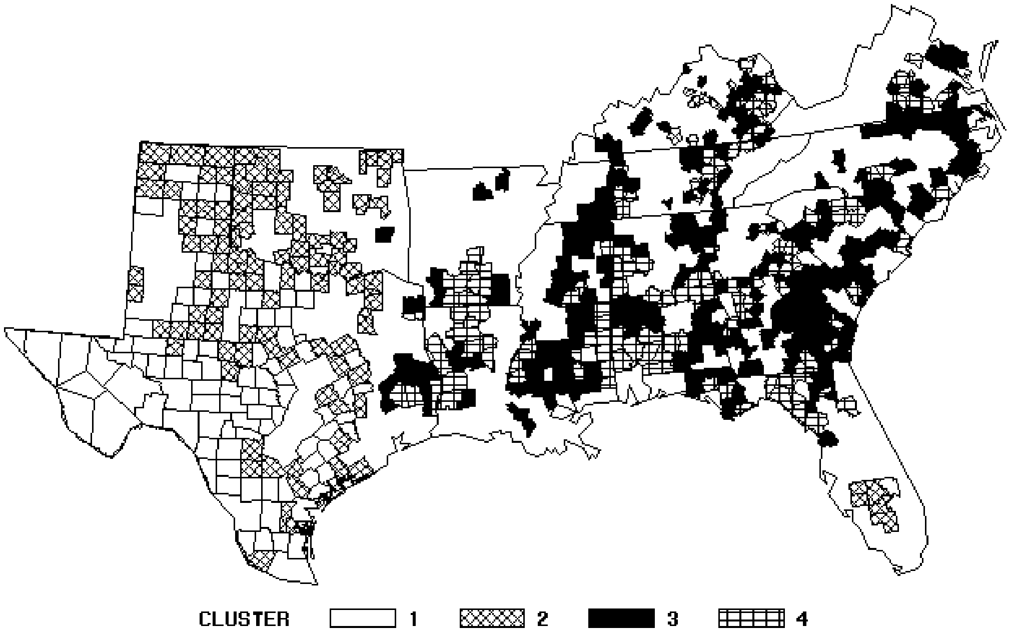
SOURCE: 1997 NORSIS, USDA-FS, Athens, GA.



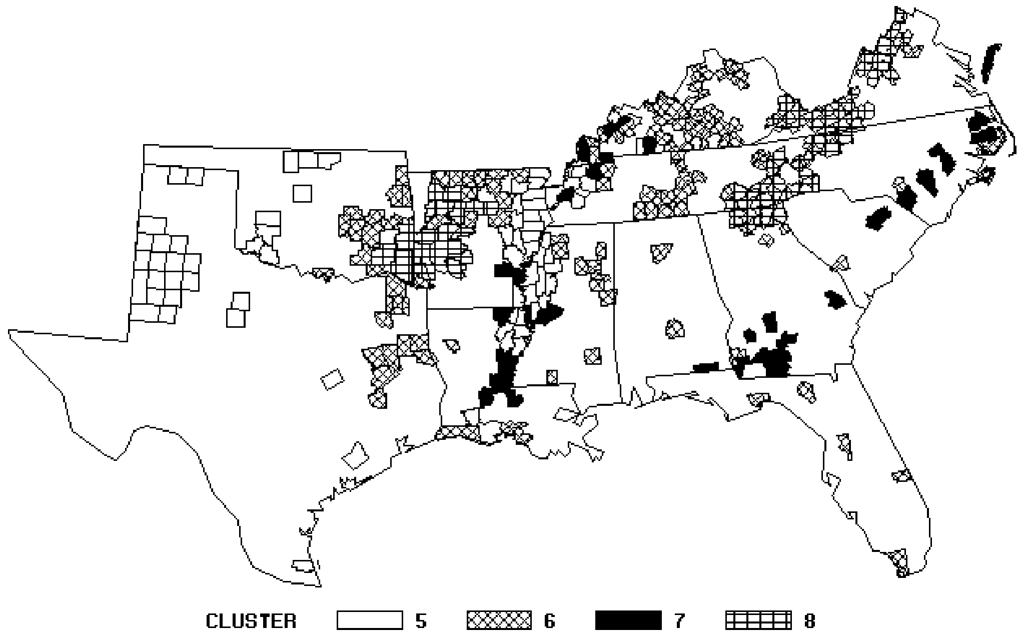
Technical Appendix Figure X.1–Counties in clusters 1 through 4, North region.



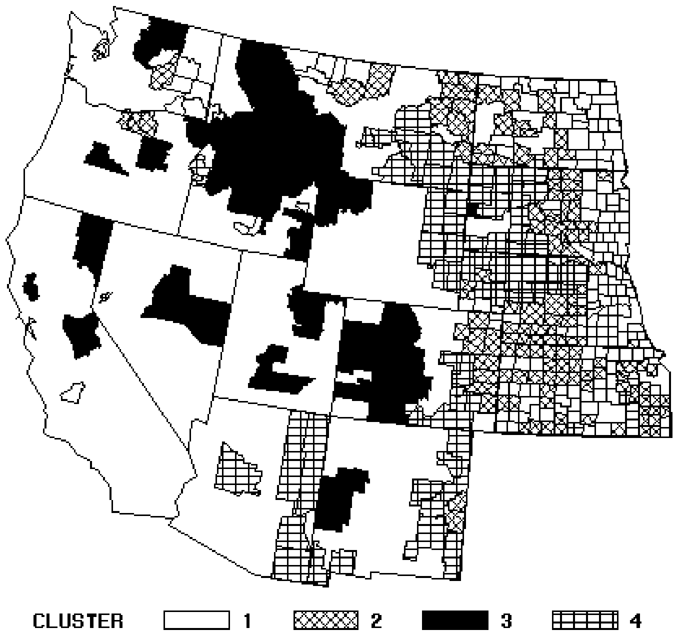
Technical Appendix Figure X.2–Counties in clusters 5 through 8, North region.



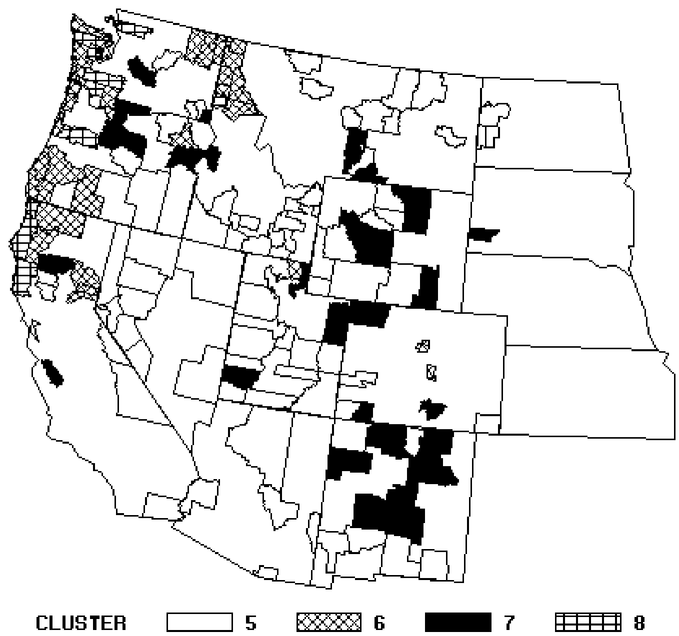
Technical Appendix Figure X.3–Counties in clusters 1 through 4, South region.



Technical Appendix Figure X.4–Counties in clusters 5 through 8, South region.



Technical Appendix Figure X.5–Counties in clusters 1 through 4, West region.



Technical Appendix Figure X.6–Counties in clusters 5 through 8, West region.